A Subjective Study of Viewing Experience for Drone Videos

Stephen Boyle, Fan Zhang and David R. Bull

Abstract

This paper presents subjective evaluation results on the viewing experience of aerial videos shot at different drone heights. A total of fifty video sequences were generated using a simulation engine, Unreal Engine 4, for two scenarios and five different shot types. Twenty human viewers were then employed to participate a subjective experiment, providing their preference opinions on viewing experience of these videos. Through the subjective test, optimal parameters of UAV height have been identified for the evaluated shot types and scenarios. These will provide recommendation of default shot parameters for drone operation in autonomous shooting and flight planning.

Index Terms

Drone cinematography, Multidrone, and optimal drone parameter.

I. Introduction

In recent years, drones have been extensively deployed as a platform for extending cinematography, supporting innovation and flexibility in shot creation, in scenarios difficult to reach by other means. After their initial adoption in film production, drones are now increasingly used in broadcasting for sports, natural history, archeology, news (e.g. natural disasters) and travelogues. They are also frequently used in user-generated content available on streaming sites.

In drone cinematography, some of the rules and heuristics that film-makers often follow to produce visually pleasing shots still work, with notable examples such as " the triangle principle" and the "rule of thirds" [1, 2]. The optimization of drone trajectories have also been researched in order to ensure camera shots meet certain cinematographic requirements and constraints. Joubert *et al.* [3] developed a system to allow drone users to visually design and preview camera shots whilst ensuring feasibility. Gebhardt *et al.* [2] developed a tool for the generation of feasible drone trajectories from sketched key-frames, each defining a time and the corresponding desired drone position.

Recently, the autonomous control of drone cameras without the need for such advanced planning has been more widely researched. Nägeli *et al.* [4] implemented an MPC (Model Predictive Control) algorithm for autonomous drone control which uses constrained optimization to ensure collision avoidance, valid drone inputs, a feasible drone state and the adherence to high level aesthetic requirements. Joubert *et al.* [5] also developed a system to autonomously control a drone to film a predefined shot sequence of one or two human subjects. However for autonomous drone systems, shot types and shot parameter defaults must be established prior to operation. There is however little research that has quantitatively investigated the relationship between viewing experience and UAV parameters, such as drone height and speed.

It is known that the experience of viewing aerial footage captured from a drone is heavily influenced by the drone and camera parameters and the relative motion between drone and target. Certain drone parameters will no-doubt lead to distracting or disorientating effects which can make the viewing experience extremely unpleasant. In order to understand viewer preferences for different shot types, these shots must be evaluated in terms of variations of the associated drone and camera parameters. This can only be done through subjective testing based on showing people (subjects) a series of shots for various scenarios, captured with differing parameter sets.

In this context, a subjective experiment has been conducted to characterise the optimal drone height for specific scenarios and shot types based on simulated video content. Compared to acquiring 'live' drone video, simulated

*The authors acknowledge funding from the European Union's Horizon 2020 (MULTIDRONE No. 731667), and the EPSRC (The Centre for Doctoral Training in Communications at University of Bristol). The authors are with the Department of Electrical and Electronic Engineering, University of Bristol, Bristol, BS8 1UB, UK.{stephen.boyle, fan.zhang, dave.bull}@bristol.ac.uk

content offers the benefits of environment flexibility, ease of generation and repeatability. It is also significantly cheaper than flying a real drone on location. The experimental results show the distribution of optimal drone height for various shot types. These results will be useful in autonomous drone operation to define default drone parameters and operational envelopes[1].

The rest of this paper is organised as follows, Section II describes how the test content was generated using a simulation engine, while the experimental methodology is presented in Section III. The subjective results are then summarised and discussed in Section IV. Finally, Section V concludes the paper and provides the future work.

II. Test Content Generation

In order to conduct a perceptual (subjective) evaluation study of this type, representative test video clips must first be acquired or generated with different drone parameters. These must not only contain relevant background and target content but must also cover the likely range of operating conditions during a shoot.

A. Using UE4 to Generate Aerial Videos

Such video content could of course be captured using actual UAVs from real scenes. This would perhaps provide the most realistic content, but is hugely time consuming in planning, obtaining permissions and shooting, and relies on either the emulation of an actual event or attendance at a real event. Apart from the resources involved, it can also be difficult to accurately control/repeat camera and drone parameters.

As an alternative to acquiring real footage, simulation engines can be employed to generate animated test video footage. In this case, test scenarios and camera/drone parameters can be carefully designed and easily changed, and there is flexibility over the choice of environment, target(s) and actions, often providing a much lower cost solution to generating large amounts of data compared to live shooting. The only possible drawback to using simulations is that the generation of specific natural-looking scenes can require significant experience.

In this work, due to the need for large amounts of test data and the requirement for accurate parameter configurations, simulation engines were used to generate video footage for subjective evaluation. By comparing multiple simulation engines including Unreal Engine 4 (UE4) [7], Unity [8] and GameMaker [9], UE4 has been adopted for this work as the simulator of choice for generating test aerial videos.

UE4 is currently the most widely used game engine, developed by Epic Games. It is relatively easy to learn and, with many developers using it, UE4 offers the largest community support. It offers a development environment which can deliver interactive virtual environments, architectural walkthroughs, training simulations, design validations and visual effects for the film industry. It is also programmable using native C++ code or visual programming with Blueprints [10].

It is noted that, for all generated simulated sequences, fixed camera and lens settings were utilised. These included a camera sensor size of 23.66mm$\times$13.3mm and focal length of 35mm (these are default parameters configured in UE4).

B. Tested Scenarios, Shot Types and Parameters

In this work, two different test scenarios were evaluated including a cycling race (three cyclists) in a countryside environment and cars racing (three cars) along a city street at night[2]. Four Unreal assets 'Country side', 'Race-Course', 'Walking Street' and 'Cycling' were obtained from the Unreal Marketplace [11] to build these scenarios. Example images are shown in Fig. 1.

Five typical UAV shot types [12] were evaluated in the conducted experiment, including STATIC, ESTABLISHING, FLYBY, CHASE and ORBIT. Their definitions are provided in Table I.

For each test shot type in both scenarios, five different height versions were generated: 1, 2, 3, 4 and 5 meters above ground level for the cycling scenario and 2, 4, 6, 8 and 10m for the RacingCar scenario. The video duration is fixed at five seconds rather than the recommended 10 seconds in ITU standard [13] based on a recent study on optimal video duration for quality assessment [14, 15]. All fifty test sequences (2 scenarios $\times$ 5 shot types $\times$ 5 height versions = 50) are acquired at 1920$\times$1080 spatial resolution, with a frame rate at 60 frames per second. Fig. 2 shows example images of five different height versions for the same Cycling-CHASE shot.

[1]As an integral part of the EU MULTIDRONE project [6], this work has been integrated into the directorial dashboard in the MULTIDRONE system.

[2]The selection of scenarios is based on the main application of the EU MULTIDRONE project [6] - live sport events.

(a) Cycling (b) RacingCar

Fig. 1: Example illustration of the tested scenarios.

TABLE I: Shot types evaluated in both scenarios

S1	**STATIC**	The drone remains stationary with no camera tracking
S2	**ESTABLISHING**	The drone moves closer to the target from the front, at a steadily decreasing altitude.
S3	**FLYBY**	The drone flies past the target (offset from the target trajectory) in a straight line, overtaking the target, with camera tracking it.
S4	**CHASE**	The drone chases the target from behind with the distance between them decreasing.
S5	**ORBIT**	The drone flies around the target in a part-circle, centred at the target.

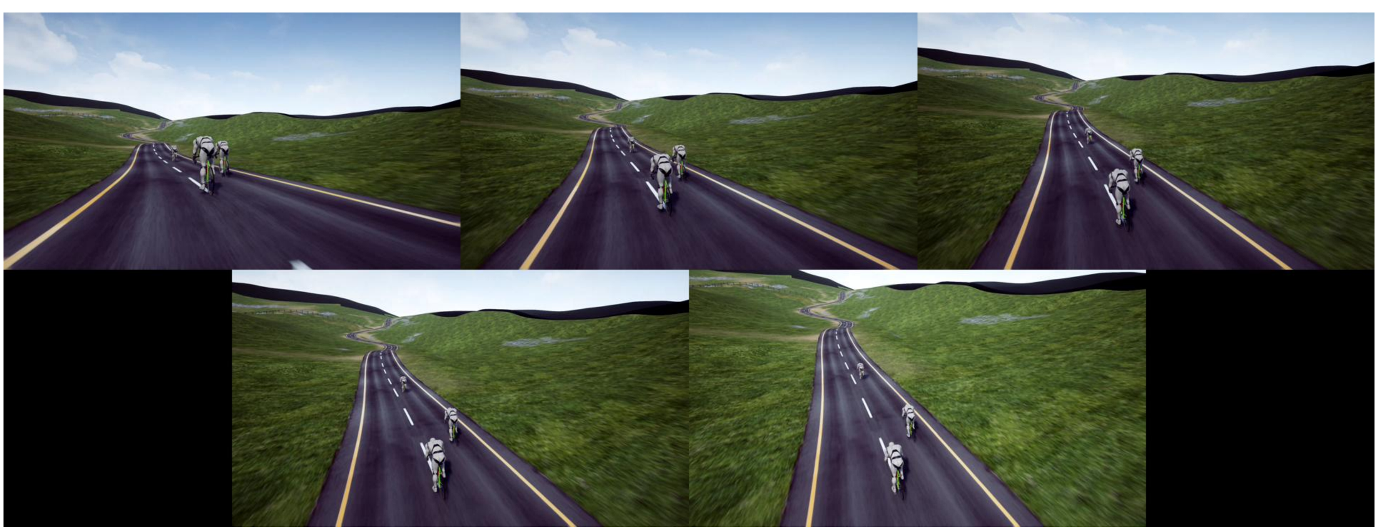


Fig. 2: Example frames from five different height versions for the Cycling CHASE shot. (Clockwise order from top left): Height values are 1, 2, 3, 4 and 5 metres.

III. Experimental Methodology

The experimental environment, test procedure, participants and data processing method employed in the experiment are described in this section.

A. Experimental Setup

The experiment was conducted in a darkened, living room style environment using a SONY KD-65ZD9 4K HDR TV, with screen size 143cm$\times$80cm. The viewing distance was set to three times the height of the TV (240cm), which is within the recommended range in ITU-R BT.500 [13]. The resolutions of the TV were configured to 1920$\times$1080 (spatial) and 60Hz (temporal). The presentation of video sequences was controlled by a Windows PC running Matlab Psychtoolbox. A second screen was employed to interact with subjects (displaying questions and collecting opinion scores).

B. Experimental Procedure

A single stimulus Absolute Category Rating (ACR) methodology was used in this experiment. In each trial, participants were shown a test sequence after viewing a 3s mid-level gray screen. Participants then had unlimited time to submit responses on the second screen, with the question, "*Please score your viewing experience 1-5, (5=Excellent, 4=Good, 3=Fair, 2=Poor and 1=Bad)*". Participants registered their answers by entering the integer numbers between 1 and 5. Before the formal test, there is a training session consisting of three training sequences (different from those in the formal test) shown to the subjects. After the whole test session, each participant was informally interviewed about their viewing experience and scoring criteria.

C. Participants and Data Processing

A total of 20 subjects (with an average age of 33) from the University of Bristol were paid to participate in the experiment including 10 males and 10 females. All of them were tested for normal or corrected-to-normal vision.

Responses from all the participants were first recorded as raw opinion scores. These were further converted to Mean Opinion Scores (MOS) for every trial (each test sequence) by taking their average value, alongside the corresponding standard error (SE). Since the experiment was designed to test the viewing experience (rather than video quality) on drone videos and the subjective results are expected to have relative high variations, outlier removal approaches have not been applied on the collected data.

IV. Results and Discussion

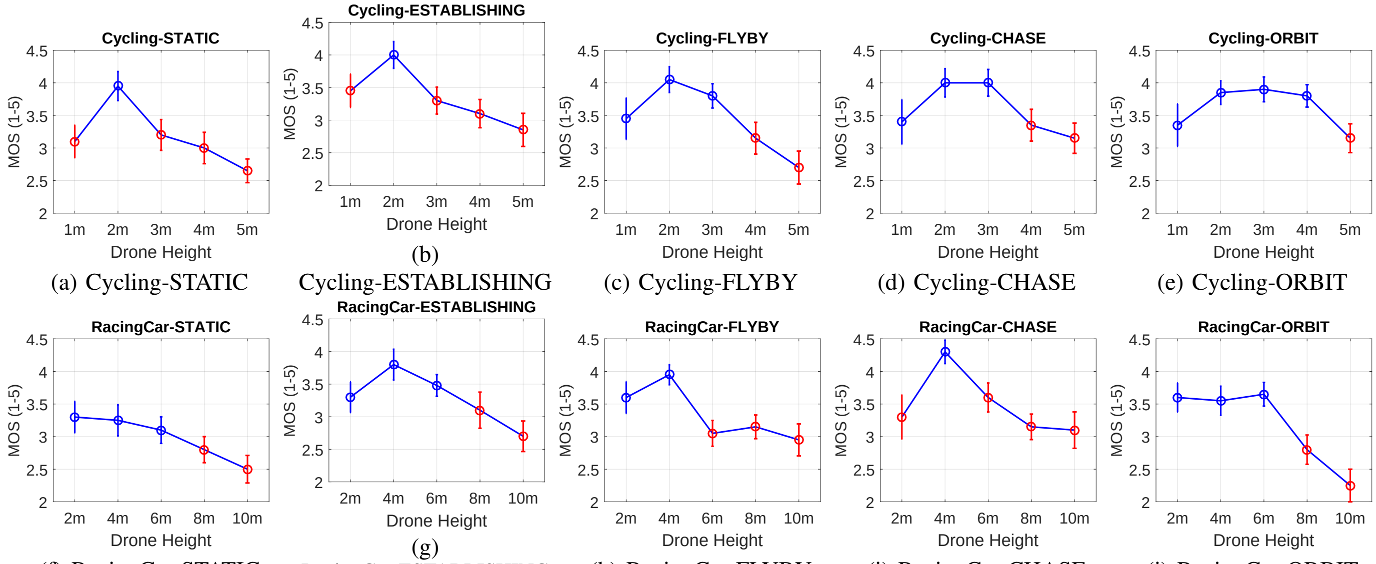
The MOS results for different height clips are presented alongside their corresponding standard errors (SE) for the tested scenarios and shot types in this section. The height parameter value with greatest MOS for each shot type is considered as optimum and compared with the other four versions for each shot type through a paired t-test at the 95% confidence interval. This indicates if the use of optimal parameter can lead to a significantly better viewing experience than the other cases.

A. Overall Test Results

The experimental results are shown in Fig. 3, in which, for most shot types, there is a statistically significant range of drone heights that give a preferential viewing experience. In some shot types (e.g. Cycling-ESTABLISHING and RacingCar-CHASE), an optimum drone height can be clearly identified. For most Cycling shots, the optimal height values are around 2 meters, which is approximately 1.4 times the cyclist height. The height value becomes greater for RacingCar shot types - 4 meters or approx. 3.4 times the car height. This may because the shapes of sport cars and cyclists are different, and the optimal drone height could also be related to object length and/or width.

B. Preference Variation

During informal interviews with participants conducted after each subjective test, it became apparent that results varied according to gender. This has been verified by the subjective opinion results, shown in Fig. 4, where the gender preferences for a selection of shot types are plotted. It can be observed that males have a slightly stronger preference to lower drone heights rather than females.



(a) Cycling-STATIC (b) Cycling-ESTABLISHING (c) Cycling-FLYBY (d) Cycling-CHASE (e) Cycling-ORBIT

(f) RacingCar-STATIC (g) RacingCar-ESTABLISHING (h) RacingCar-FLYBY (i) RacingCar-CHASE (j) RacingCar-ORBIT

Fig. 3: Results of experiment testing the effect of drone (camera) height. Here the error bar represents the standard error, and the red points stand for the MOS of test parameters which are significantly (through a paired t-test at the 95% confidence interval) lower than each best case.

C. *The Application of the Optimal Parameters*

The primary objective of this work was to determine default (optimal) height parameters for typical shot types and scenarios in potential autonomous flight and shooting. It is noted that all the optimal parameters obtained above are based on specific camera and lens settings during simulation. In practice, in order to obtain the same Field of View (FOV) or framing, the working distance (WD), e.g. height of the drone, can be converted for the actually used camera settings (Sensor Size SS and Focal Length FL). This conversion is based on the basic calculation formula of Field of View (FOV) [16], which is shown in Fig. 5 and equation (1).

$$\mathrm{FL} \times \mathrm{FOV} = \mathrm{SS} \times \mathrm{WD} \tag{1}$$

The actual working distance $\mathrm{WD}_{\mathrm{act}}$ can be calculated by:

$$\mathrm{WD}_{\mathrm{act}} = \frac{\mathrm{SS}_{\mathrm{ref}}}{\mathrm{SS}_{\mathrm{act}}} \times \frac{\mathrm{FL}_{\mathrm{act}}}{\mathrm{FL}_{\mathrm{ref}}} \times \mathrm{WD}_{\mathrm{ref}} \tag{2}$$

in which $\mathrm{SS}_{\mathrm{ref}}$ and $\mathrm{FL}_{\mathrm{ref}}$ are the camera parameters used for generating the simulation videos that are given in Section II-A, while $\mathrm{WD}_{\mathrm{ref}}$ is the recommended working distance, e.g. the optimal drone heights determined above. $\mathrm{SS}_{\mathrm{act}}$ and $\mathrm{FL}_{\mathrm{act}}$ are actual camera parameters used in practice.

It is noted that in real media production using drones, directorial decision making will be key in determining optimum working distance to obtain meaningful shots. The results generated in this work are however valuable in providing default parameters for the tested scenarios and shot types.

V. Conclusion

In this paper, the results of a subjective study are presented on the relationship between drone height and viewing experience. The test scenarios and shot types were carefully designed and generated using a simulation engine UE4. The subjective results show consistent preference of UAV heights from employed participants. Future work will focus on the influence of drone speed and the relationship between these optimal parameters and target object sizes.

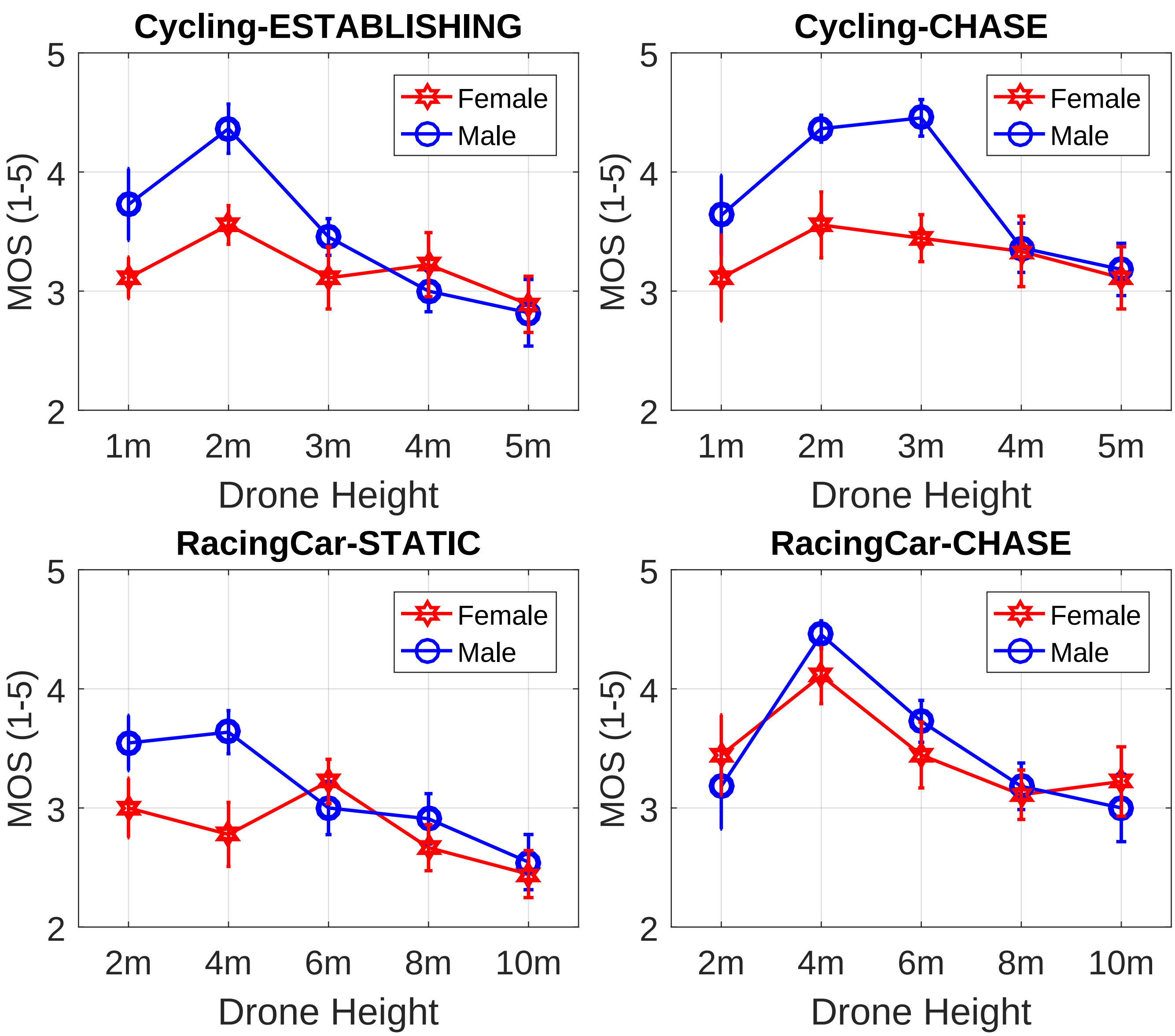


Fig. 4: Results of gender preferences on selected scenarios ans shot types. In each sub-figure, the blue points stand for the MOS scores for male participants only, while the red points represent those for female subjects. The error bars indicate the standard errors for the MOS scores.

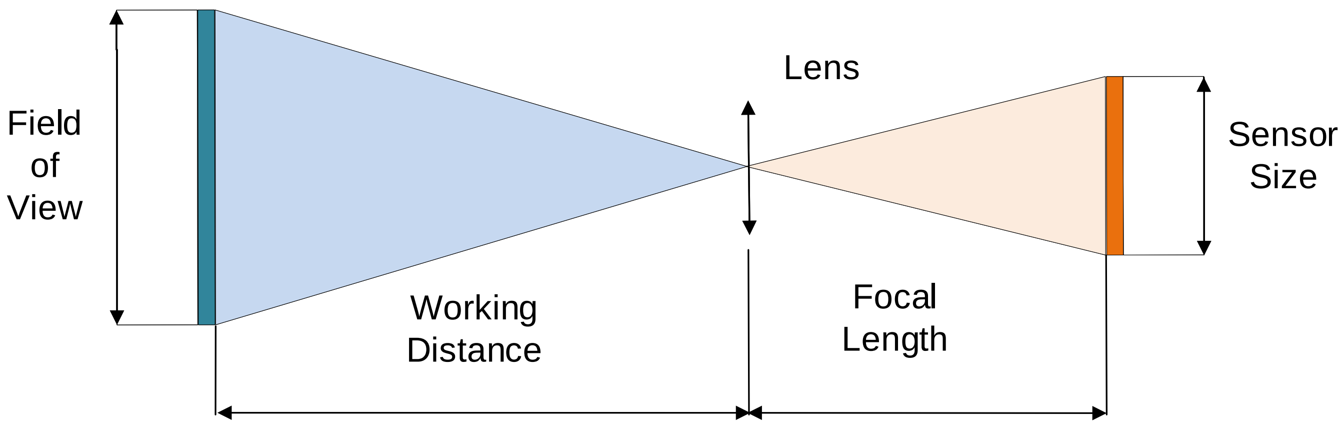


Fig. 5: Calculation of FOV.

VI. References

[1] D. Arijon, *Grammar of the Film Language*. London, UK: Focal Press, 1976.

[2] C. Gebhardt, B. Hepp, T.Nägeli, S.Stevšić, and O.Hilliges, "AirWays: Optimization-Based Planning of Quadrotor Trajectories according to High-Level User Goals," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, May 2016, pp. 2508–2519.

[3] N. Joubert, M. Roberts, A.Truong, F. Berthouzoz, and P. Hanrahan, "An Interactive Tool for Designing Quadrotor Camera Shots," *ACM Transactions on Graphics*, vol. 34, no. 6, pp. 1–11, 2015.

[4] T. Nägeli, J. Alonso-Mora, A. Domahidi, D. Rus, and O. Hilliges, "Real-time Motion Planning for Aerial Videography with Dynamic Obstacle Avoidance and Viewpoint Optimization," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1696–1703, 2017.

[5] N. Joubert, J. E, D. Goldman, F. Berthouzoz, and M. Roberts, "Towards a Drone Cinematographer: Guiding Quadrotor Cameras using Visual Composition Principles," 2016. [Online]. Available: https://arxiv.org/abs/1610.01691

[6] "Multidrone – using multiple drones for media production," 2017. [Online]. Available: https://multidrone.eu/

[7] Epic Games, "Unreal engine." [Online]. Available: https://www.unrealengine.com

[8] Unity Technologies, "Unity." [Online]. Available: https://unity3d.com/

[9] YoYo Games , "Gamemaker." [Online]. Available: https://www.yoyogames.com/gamemaker

[10] Epic Games, "Unreal Engine 4 Documentation." [Online]. Available: https://docs.unrealengine.com/latest/INT/

[11] ——, "UE4 Marketplace." [Online]. Available: https://www.unrealengine.com/marketplace/

[12] C. Smith, *The Photographer's Guide to Drones*. Rocky Nook, Inc., 2016.

[13] *Methodology for the subjective assessment of the quality of television pictures*, ITU-R Std. Recommendation ITU-R BT.500-13, 2012.

[14] F. Mercer Moss, K. Wang, F. Zhang, R. Baddeley, and D. R. Bull, "On the optimal presentation duration for subjective video quality assessment," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 26, no. 11, pp. 1977–1987, 2016.

[15] F. Mercer Moss, C.-T. Yeh, F. Zhang, R. Baddeley, and D. R. Bull, "Support for reduced presentation durations in subjective video quality assessment," *Signal Processing: Image Communication*, vol. 48, pp. 38–49, 2016.

[16] G. C. Holst, *Testing and evaluation of infrared imaging systems*. JCD Pub., 1998.